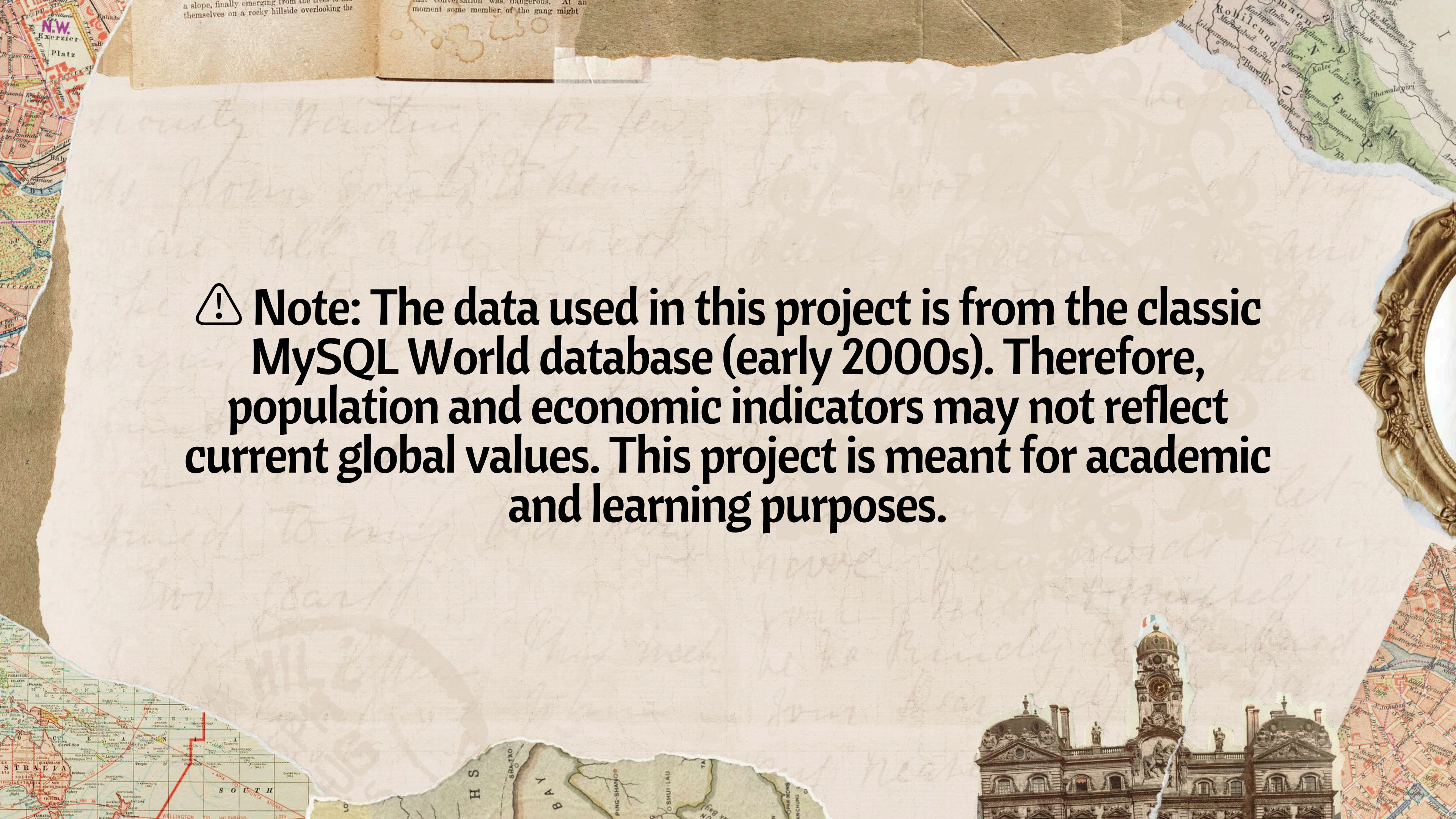


SQL Project: Analyzing Global Data from the World Database

Using MySQL & PowerBI on Country, City, CountryLanguage Tables

By - Dev . N . Zaveri



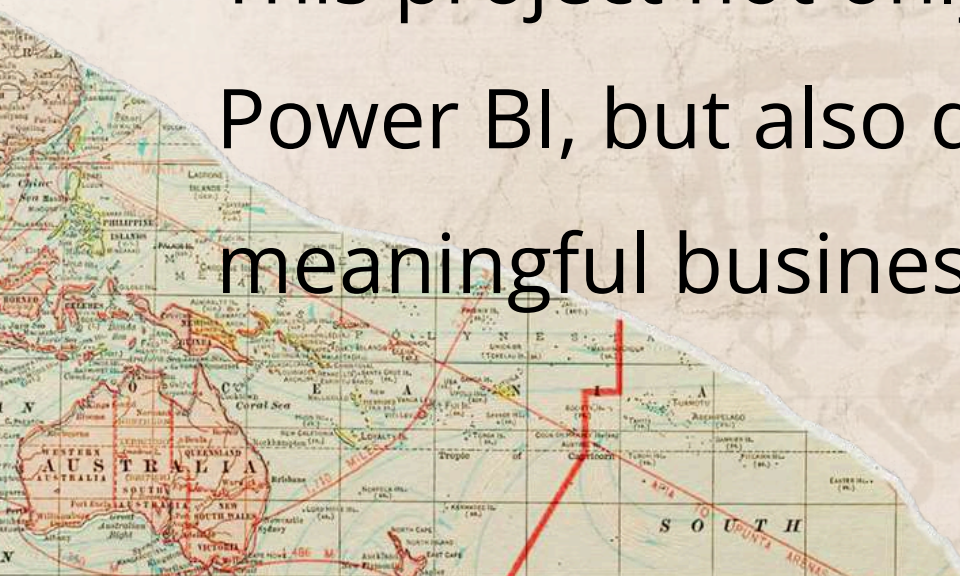
⚠ Note: The data used in this project is from the classic MySQL World database (early 2000s). Therefore, population and economic indicators may not reflect current global values. This project is meant for academic and learning purposes.

Introduction

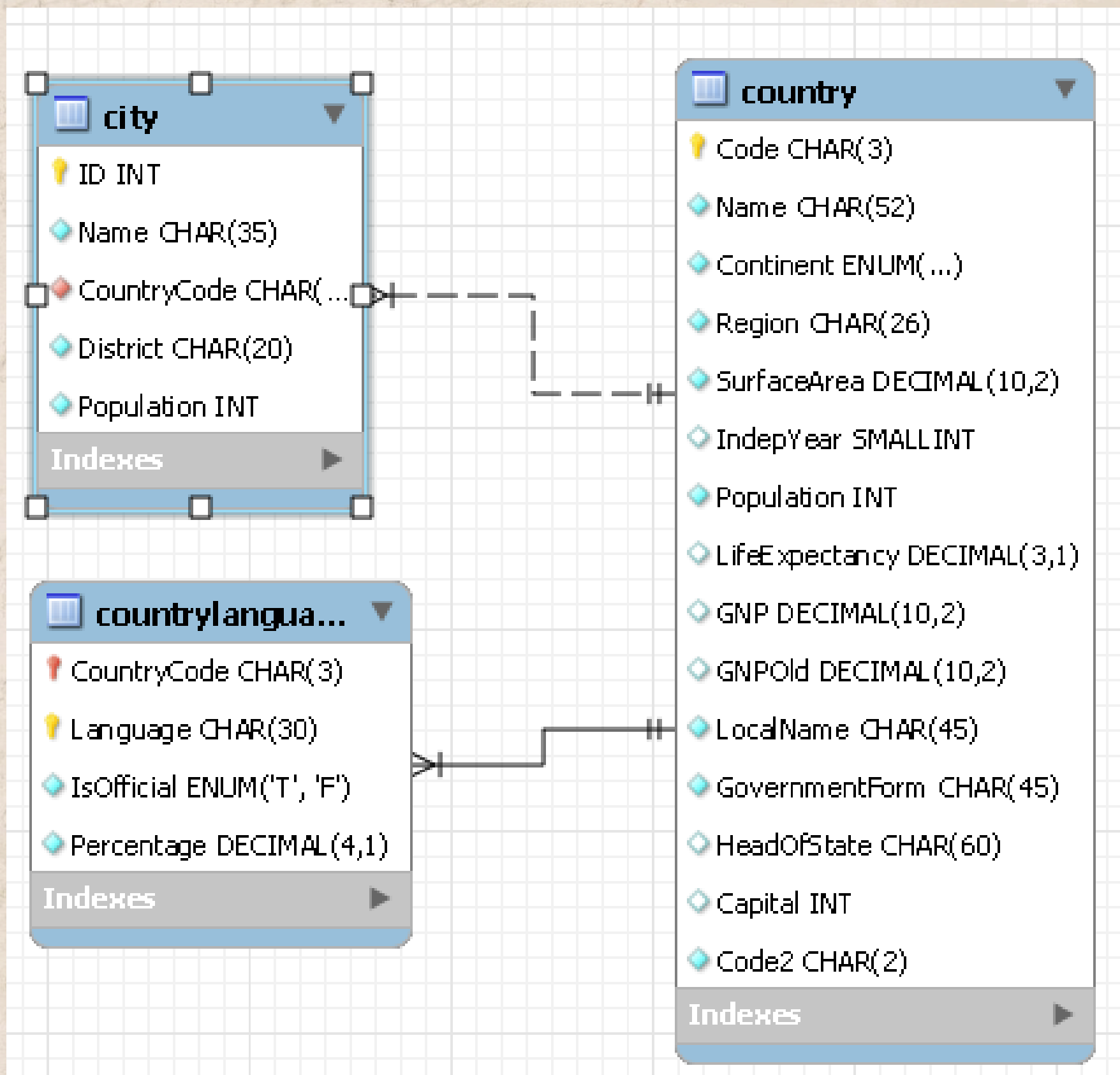
This project explores and analyzes global demographic and linguistic data using the classic World Database provided in MySQL. The database includes three interconnected tables — country, city, and countrylanguage — which provide detailed information about nations, their cities, and the languages spoken.

Using structured SQL queries, I extracted key insights such as global population trends, dominant languages, and city-level demographics. These insights were then brought to life through interactive dashboards in Power BI, enabling clear and impactful data storytelling.

This project not only strengthened my skills in SQL, relational databases, and Power BI, but also deepened my understanding of how to transform raw data into meaningful business insights.



Database Schema



🧩 Schema Explanation:

The World Database is structured into three main tables:

- **country:** Contains information about each country, including its name, population, surface area, government form, and more.
- **city:** Lists cities along with their name, population, and the CountryCode to indicate which country they belong to.
- **countrylanguage:** Details the languages spoken in each country, with information about whether the language is official and the percentage of speakers.

The tables are connected by foreign keys:

- `city.CountryCode` → `country.Code`
- `countrylanguage.CountryCode` → `country.Code`

These relationships allow us to join the tables and answer complex questions involving cities, countries, and the languages spoken.



About the Questions Answered

In this SQL project, I explored various aspects of global data using real-world questions that demonstrate my understanding of database relationships, data aggregation, and analytical thinking.

The questions were designed to cover a wide range of insights, including:

- City-Level Analysis (e.g., most populous cities, city distribution by country)
- Country-Level Analysis (e.g., countries with highest GNP, population contribution to continents)
- Language Usage Patterns (e.g., most spoken and official languages)
- Continental Comparisons (e.g., number of countries per continent, average life expectancy).

These questions helped me practice and apply SQL concepts such as:

- JOINS to combine related data from multiple tables
- GROUP BY, ORDER BY, and LIMIT for grouping, sorting, and selecting top results
- Aggregate functions like COUNT(), AVG(), SUM()
- Subqueries for calculating derived values such as averages or totals

This diverse set of questions helped transform raw data into meaningful insights and laid the foundation for data visualization using Power BI.

Q1:Query the top 10 cities by population from the city table?

Query:

```
SELECT  
    name, population  
FROM  
    city  
ORDER BY population DESC  
LIMIT 10;
```

Result:

	name	population
►	Mumbai (Bombay)	10500000
	Seoul	9981619
	São Paulo	9968485
	Shanghai	9696300
	Jakarta	9604900
	Karachi	9269265
	Istanbul	8787958
	Ciudad de México	8591309
	Moscow	8389200
	New York	8008278

Q2:How many countries exist in each continent?

Query:

```
• SELECT
    continent, COUNT(name) AS country_count
FROM
    country
GROUP BY continent;
```

Result:

continent	country_count
North America	37
Asia	51
Africa	58
Europe	46
South America	14
Oceania	28
Antarctica	5

Q3: Show the average life expectancy grouped by continent?

Query:

```
SELECT
    continent,
    ROUND(AVG(Lifeexpectancy), 2) AS avg_lifeexpectancy
FROM
    country
GROUP BY continent;
```

Result:

continent	avg_lifeexpectancy
North America	72.99
Asia	67.44
Africa	52.57
Europe	75.15
South America	70.95
Oceania	69.72
Antarctica	NULL

Q4:List top 5 countries whose GNP is higher than the average GNP?

Query:

```
SELECT  
    name, GNP  
FROM  
    country  
WHERE GNP > (SELECT AVG(GNP) FROM country)  
ORDER BY GNP DESC  
LIMIT 5;
```

Result:

name	GNP
United States	8510700.00
Japan	3787042.00
Germany	2133367.00
France	1424285.00
United Kingdom	1378330.00

Q5: Find the top 5 languages with the highest average percentage of use?

Query:

```
SELECT
    language, AVG(percentage) AS avg_percentage
FROM
    countrylanguage
GROUP BY language
ORDER BY avg_percentage DESC
LIMIT 5;
```

Result:

	language	avg_percentage
►	Kirundi	98%
	Marshallese	97%
	Maltese	96%
	Icelandic	96%
	Bajan	95%

Q6: List languages marked as IsOfficial = 'T' and how many countries use them?

Query:

```
• SELECT  
    COUNT(DISTINCT countrycode) AS country_count,  
    COUNT(DISTINCT language) AS language_count  
FROM  
    countrylanguage  
WHERE  
    IsOfficial = 'T';
```

Result:

	country_count	language_count
▶	190	102

Q7: List the top 10 most populated cities along with their respective country names?

Query:

```
SELECT
    city.name AS city_name,
    city.population AS Total_population,
    country.name AS country_name
FROM
    city
    JOIN
    country ON city.countrycode = country.Code
ORDER BY city.population DESC
LIMIT 10;
```

Result:

city_name	Total_population	country_name
Mumbai (Bombay)	10500000	India
Seoul	9981619	South Korea
São Paulo	9968485	Brazil
Shanghai	9696300	China
Jakarta	9604900	Indonesia
Karachi	9269265	Pakistan
Istanbul	8787958	Turkey
Ciudad de México	8591309	Mexico
Moscow	8389200	Russian Federation
New York	8008278	United States

Q8: Calculate what % of the continent's population each country holds?

Query:

```
SELECT
    country.Name AS country_name,
    country.Continent,
    SUM(city.Population) AS country_population,
    continent_pop.total_population,
    CONCAT(ROUND(SUM(city.Population) / continent_pop.total_population * 100, 2), '%') AS percentage_share
FROM country
JOIN city ON city.CountryCode = country.Code
JOIN (
    SELECT country.Continent, SUM(city.Population) AS total_population
    FROM country
    JOIN city ON city.CountryCode = country.Code
    GROUP BY country.Continent
) AS continent_pop ON country.Continent = continent_pop.Continent
GROUP BY country.Name, country.Continent, continent_pop.total_population
ORDER BY ROUND(SUM(city.Population) / continent_pop.total_population * 100, 2) DESC
LIMIT 10;
```


Q8 Result:

	country_name	Continent	country_population	total_population	percentage_share
►	Australia	Oceania	11313666	13886149	81.47%
	Brazil	South America	85876862	172037859	49.92%
	United States	North America	78625774	168250381	46.73%
	Mexico	North America	59752521	168250381	35.51%
	Russian Federation	Europe	69150700	241942813	28.58%
	China	Asia	175953614	697604103	25.22%
	India	Asia	123298526	697604103	17.67%
	Egypt	Africa	20083079	135838579	14.78%
	New Zealand	Oceania	1847600	13886149	13.31%
	Nigeria	Africa	17366900	135838579	12.78%

Q9:List the top 5 countries with the highest number of cities having a population over 1 million?

Query:

```
SELECT
    country.name AS country_name, COUNT(city.name) AS city_name
FROM
    country
    JOIN
    city ON country.Code = city.CountryCode
WHERE
    city.population > 1000000
GROUP BY country_name
ORDER BY city_name DESC
LIMIT 5;
```

Result:

	country_name	city_name
►	China	35
	India	18
	Brazil	13
	Russian Federation	12
	Japan	11

Q10: Which continent has the highest average city population?

Query:

```
SELECT
    country.Continent AS Continent,
    ROUND(AVG(city.Population), 2) AS avg_population
FROM
    country
    JOIN
    city ON country.code = city.CountryCode
GROUP BY Continent
ORDER BY avg_population DESC;
```

Result:

	Continent	avg_population
▶	Asia	395019.31
	Africa	371143.66
	South America	366038.00
	North America	289587.57
	Europe	287684.68
	Oceania	252475.44

Q11:List the top 7 most populated cities along with their country name and the most spoken official language in that country?

Query:

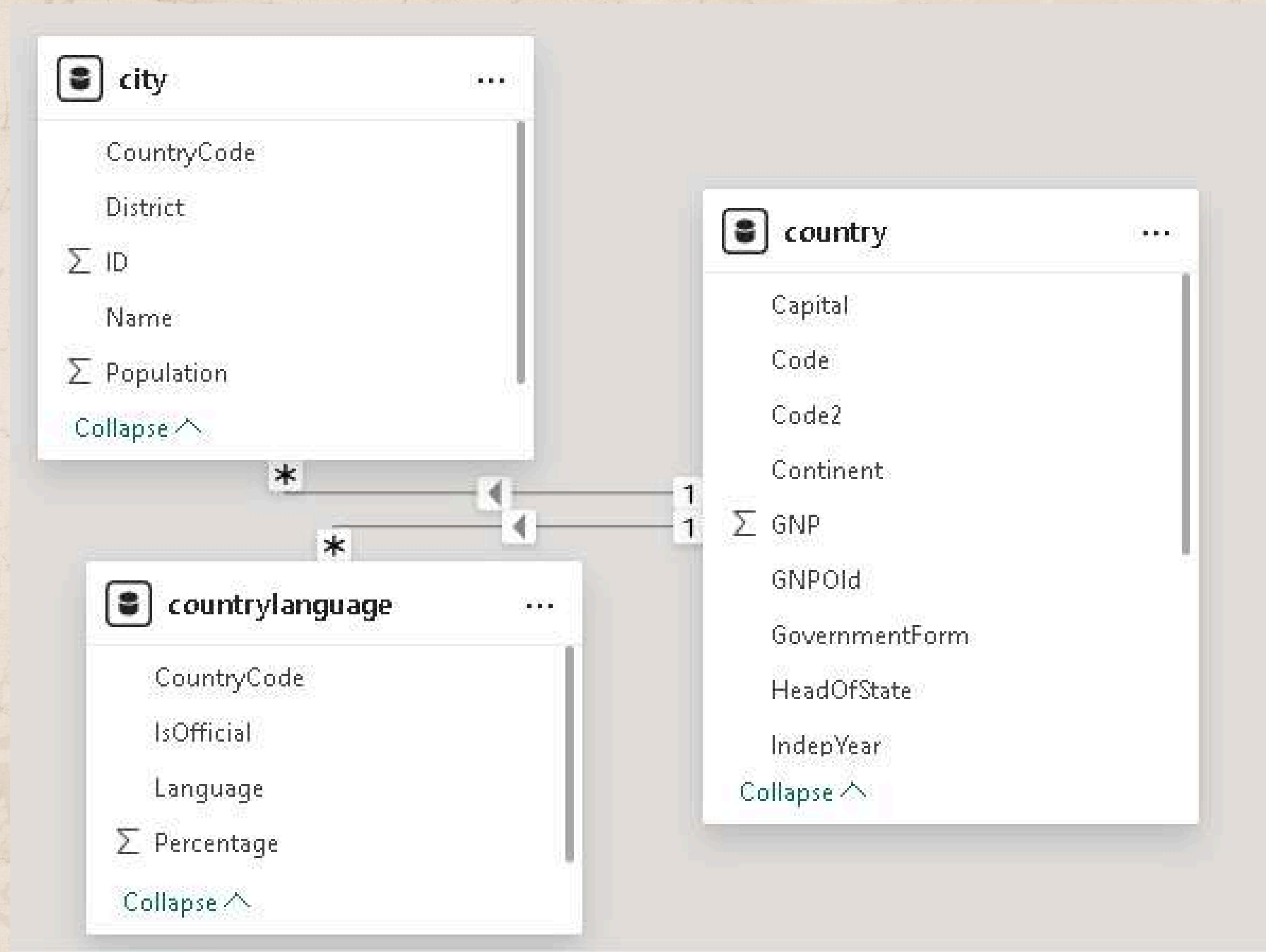
```
SELECT
    city.name AS City_name,
    city.population AS population,
    country.name AS Country_name,
    countrylanguage.Language AS official_language
FROM
    city
    JOIN
    country ON city.CountryCode = country.Code
    JOIN
    countrylanguage ON Country.Code = countrylanguage.CountryCode
WHERE
    countrylanguage.IsOfficial = 'T'
ORDER BY city.population DESC
LIMIT 7;
```

Result:

	City_name	population	Country_name	official_language
►	Mumbai (Bombay)	10500000	India	Hindi
	Seoul	9981619	South Korea	Korean
	São Paulo	9968485	Brazil	Portuguese
	Shanghai	9696300	China	Chinese
	Jakarta	9604900	Indonesia	Malay
	Karachi	9269265	Pakistan	Urdu
	Istanbul	8787958	Turkey	Turkish

From SQL to Insights Visualizing Global Data with Power BI

PowerBI Model view



Power BI Relationships Explained

Relationships Between City, Country, and CountryLanguage Tables in Power BI

In this model, three tables from separate Excel files are connected using a common key:

- **City[CountryCode] is related to Country[Code]**
- **CountryLanguage[CountryCode] is also related to Country[Code]**

These relationships form a centralized model where Country acts as the core table, connecting to both City and CountryLanguage.

This allows us to analyze data such as:

- **Cities by country**
- **Languages spoken in each country**
- **Aggregated insights like number of cities per country or top spoken languages globally**



Global Insights Dashboard – World SQL Database

CONTINENT

- ☐ South America
- ☐ Oceania
- ☐ North America
- ☐ Europe
- ☐ Asia
- ☐ Antarctica
- ☐ Africa

Official Language

- ☒ Afrikaans
- ☒ Ainar
- ☒ Albanian
- ☒ Arabic
- ☒ Armenian
- ☒ Azerbaijani
- ☒ Belarusian
- ☒ Bengali
- ☒ Bislama
- ☒ Bulgarian
- ☒ Burmese
- ☒ Catalan
- ☒ Chamorro
- ☒ Chichewa

239

Total Countries

4001

Total Cities

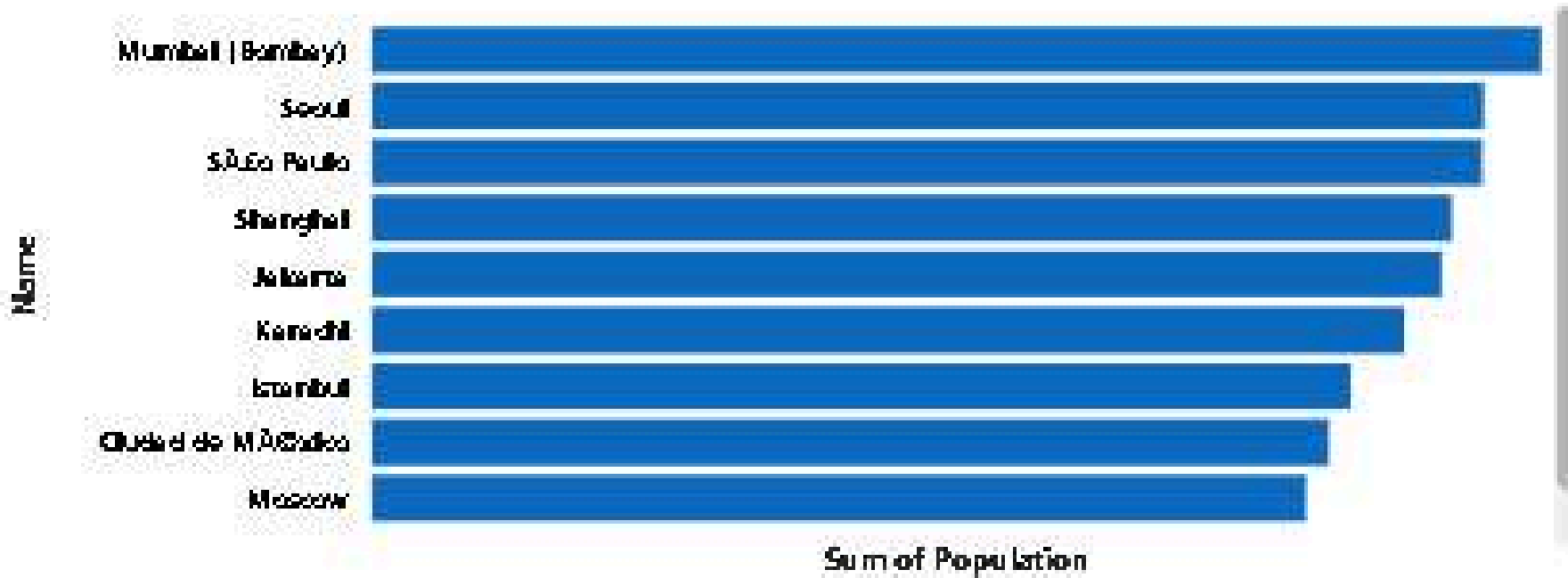
1,429,559,884

Total Population

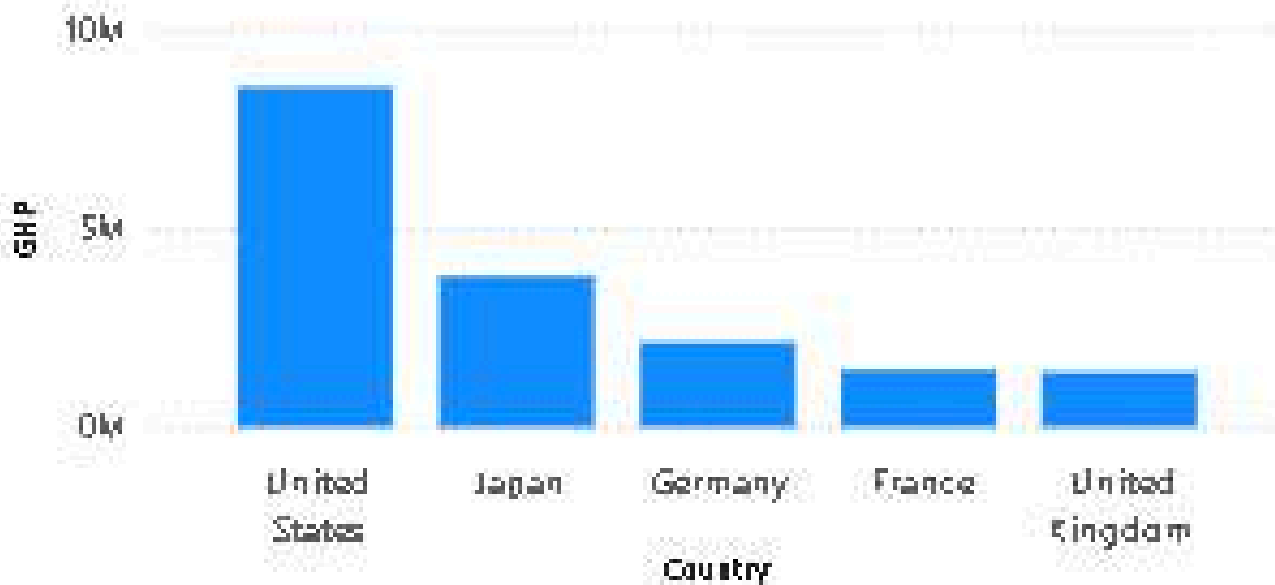
7

Total Continents

Top 10 most populated cities(Based on SQL)



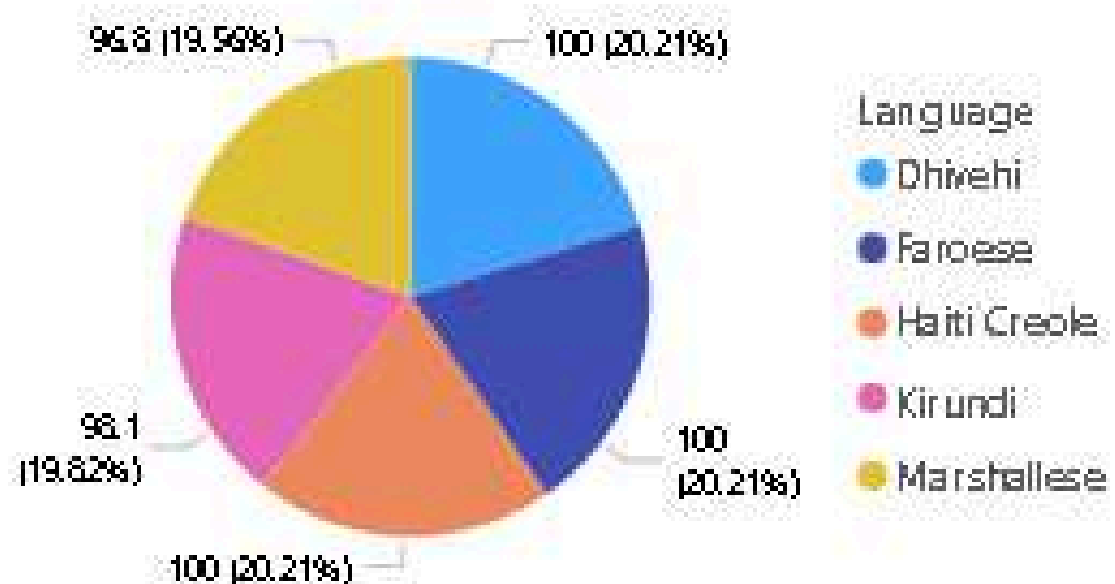
Top 5 countries with GNP higher than the average



Global Population by Country



Top 5 Languages by Avg Percentage of Use





Dashboard Overview – “Global City & Country Insights”

This interactive dashboard was created using Power BI and data extracted via SQL queries from the MySQL World Database. It presents a high-level view of global population, cities, countries, languages, and economic indicators.



Key Features:

- ✓ Top 10 Most Populated Cities – Based on SQL output, highlighting cities with the highest populations.
- ✓ Top 5 Countries by GNP (Above Average) – Showcasing economic strength using aggregated GNP data.
- ✓ Pie Chart for Language Distribution – Displays the most widely used official languages by percentage of usage.
- ✓ Interactive Map – Visualizes global population spread with bubble sizes representing city populations.
- ✓ Card KPIs – Quick glance metrics: total countries, cities, continents, and global population.
- ✓ Slicers – Allow filtering by continent and official language for deeper exploration.



Purpose & Value:

This dashboard makes it easy to analyze global patterns, compare cities and countries, and identify key insights related to demographics and languages. It’s designed to be visually simple yet rich in insights.



Thank you very much!

Presented by Dev .N. Zaveri
SQL + Power BI Project